GI5321A

Crystal Structure of *E. coli* GDP-Fucose Synthetase (and Complexes Thereof) and Methods of Idnetifying Agonists and Antagonists Using Same

This application claims priority from U.S. application Ser. No. 60/096,452, filed August 13, 1998.

Background of the Invention

Fucose is found widely distributed in the complex carbohydrates and glycoconjugates of bacteria, plants, and animals. In these organisms it plays diverse roles, ranging from its involvement in nodulation in *Azorhizobium* [1] to development of shoots in *Arabidopsis* [2] to adhesion of leukocytes to activated endothelia in humans as part of the selectin ligand [3]. In humans a defect in GDP-fucose biosynthesis is responsible for the immune disorder Leukocyte Adhesion Deficiency type II [4, 5, 6]. Fucose is added to these glycoconjugates by specific transferases that use GDP-fucose as the sugar donor. GDP-fucose in turn is synthesized primarily from GDP-mannose in a three-step reaction involving two enzymes as shown in Figure 1. The first step is the oxidation at C4 of the mannose ring and subsequent reduction at C6. This is carried out by a $NADP^+$ dependent enzyme, GDP-mannose 4,6 dehydratase (GMD) [7, 8, 9]. The next two steps of the reaction, the epimerization at C3 and C5 of the mannose ring and the

subsequent NADPH dependent reduction at C4 to yield GDP-fucose, are carried out by a single dual function enzyme, GDP-fucose synthetase (GFS) [9, 10, 11]. In *E. coli* this enzyme is encoded by the *fcl* gene, previously known as *wca*G [12, 13]. It is in these final two steps that GDP-fucose biosynthesis differs from synthesis of other deoxy sugars derived from dTDP-glucose and CDP-glucose. In the latter pathways, separate epimerase and reductase enzymes encoded by independent genes perform the roles of the dual function epimerase-reductase of the GDP-fucose pathway (reviewed in [14]).

The human homologue of GFS has recently been identified as the FX protein [11]. As with the *E. coli* enzyme it is a homodimer that binds NADP(H) and catalyzes both the epimerization and reduction of GDP- 4-keto, 6-deoxy-mannose. Human GFS has 29% identity to the *E. coli* protein. More distantly related to both the human and *E. coli* enzymes is UDP-galactose-4-epimerase (GalE), which catalyzes the reversible epimerization of UDP-glucose to UDP-galactose. Essential to catalysis is a tightly bound NAD^+ that is reduced and then oxidized during the catalytic cycle. UDP-galactose 4-epimerase is a member of the short chain family of dehydrogenase/reductases (SDR) (reviewed in [15]). This family of enzymes catalyzes a diverse set of enzymatic reactions spanning 5 E.C. classes using a conserved set of active site residues including a Ser-Tyr-Lys catalytic triad.

It would, therefore, be desirable to determine the structure of *E. coli* GDP-fucose synthetase in order to facilitate the identification and development of agonists and antagonists of GFS enzyme activity in humans and other species.

Summary of the Invention

We have determined the structure of GDP-fucose synthetase from *E. coli* at 2.2Å resolution. The structure of GDP-fucose synthetase is closely related to that of UDP-galactose 4-epimerase and more distantly to other members of the short chain dehydrogenase/reductase family. We have also determined the structures of the binary complexes of GDP-fucose synthetase with its substrate NADPH and its product $NADP^+$. The nicotinamide cofactors bind in the *syn* or *anti* conformations, respectively.

GDP-fucose synthetase binds its substrate, NADPH, in the proper orientation (*syn*) to transfer the pro-*S* hydride. We have observed a single binding site in GDP-fucose synthetase for the second substrate, GDP-4-keto, 6-deoxy-mannose. This implies that both the epimerization and reduction reactions occur at the same site on the enzyme. As for all members of the short-chain family of dehydrogenase/reductases, GDP-fucose synthetase retains the Ser-Tyr-Lys catalytic triad. We propose that this catalytic triad functions in a mechanistically equivalent manner in both the epimerization and reduction reactions. Additionally, the x-ray structure has allowed us to identify other residues potentially substrate binding and catalysis.

The present invention provides for crystalline GFS. Preferably, the GFS is *E. coli* GFS, although GFS from other species are also included within the invention. In certain embodiments, the GFS is recombinant GFS and/or comprises the mature sequence of naturally-occurring GFS.

Other embodiments provide for a crystalline composition comprising GFS is association with a second chemical species. Preferably, the second chemical species is selected from the group consisting of NADPH, NADP+ and a potential inhibitor of GFS activity.

Yet other embodiments provide for a model of the structure of GFS comprising a data set embodying the structure of GFS. Preferably, such data set was determined by crystallographic analysis of GFS, including possibly by NMR analysis. In certain embodiments, the data set embodies a portion of the structure of GFS, including without limitation the active site of GFS.

Any available method may be used to construct such model from the crystallographic and/or NMR data disclosed herein or obtained from independent analysis of crystalline GFS. Such a model can be constructed from available analytical data points using known software packages such as HKL, MOSFILM, XDS, CCP4, SHARP, PHASES, HEAVY, XPLOR, TNT, NMRCOMPASS, NMRPIPE, DIANA, NMRDRAW, FELIX, VNMR, MADIGRAS, QUANTA, BUSTER, SOLVE, O, FRODO, RASMOL, and CHAIN. The model constructed from these data can then be visualized using available systems, including, for example, Silicon Graphics, Evans and Sutherland, SUN, Hewlett Packard, Apple Macintosh, DEC, IBM, and Compaq. The present invention also provides for a computer system which comprises the model of the invention and hardware used for construction, processing and/or visualization of the model of the invention.

Further embodiments provide a computer system comprising computer hardware and the model of the present invention.

Methods are also provided for identifying a species which is an agonist or antagonist of GFS activity or binding comprising: (a) providing the model of the present invention, (b) studying the interaction of candidate species with such model, and (c) selecting a species which is predicted to act as said agonist or antagonist. Species identified in accordance with such methods are also provided.

Other embodiments provide: (1) a process of identifying a substance that inhibits GFS activity or binding comprising determining the interaction between a candidate substance and a model of the structure of GFS, or (2) a process of

identifying a substance that mimics GFS activity or binding comprising determining the interaction between a candidate substance and a model of the structure of GFS. Substances identified in accordance with such processes are also provided.

The study of the interaction of the candidate species with the model can be performed using available software platforms, including QUANTA, RASMOL, O, CHAIN, FRODO, INSIGHT, DOCK, MCSS/HOOK, CHARMM, LEAPFROG, CAVEAT(UC Berkley), CAVEAT(MSI), MODELLER, CATALYST, and ISIS.

Other embodiments provide a method of identifying inhibitors of GFS activity by rational drug design comprising: (a) designing a potential inhibitor that will form non-covalent bonds with one or more amino acids in the GFS sequence based upon the crystal structure co-ordinates of GFS; (b) synthesizing the inhibitor; and (c) determining whether the potential inhibitor inhibits the activity of GFS. In other preferred embodiments, the inhibitor is designed to interact with one or more amino acids selected from the group consisting of Arg12, Met14, Val15, Arg36, Asn40, Leu41, Ala63, Ile86, Gly106, Ser107, Ser108, Cys109, Tyr136, Lys140, Asn165, Leu166, His179, Val180, Leu184, Val201, Trp202, Arg209, and Lys283.

Agonists and antagonists identified by such methods are also provided.

A process is also provided of identifying a substance that inhibits human FX protein activity or binding comprising determining the interaction between a candidate substance and a model of the structure of GFS of the present invention.

Other embodiments provide for a method of identifying inhibitors of human FX protein activity by rational drug design comprising:

(a) designing a potential inhibitor that will form non-covalent bonds with one or more amino acids in the GFS sequence based upon the crystal structure co-ordinates of crystalline GFS of the present invention;

(b) synthesizing the inhibitor; and

(c) determining whether the potential inhibitor inhibits the activity of human FX protein.

In preferred embodiments, the inhibitor is designed to interact with one or more amino acids in the GFS sequence selected from the group consisting of Arg12, Met14, Val15, Arg36, Asn40, Leu41, Ala63, Ile86, Gly106, Ser107, Ser108, Cys109, Tyr136, Lys140, Asn165, Leu166, His179, Val180, Leu184, Val201, Trp202, Arg209, and Lys283.

Agonists and antagonists identified by such methods are also provided.

Brief Description of the Figures

Figure 1: The GDP-fucose biosynthetic pathway. The enzymes catalyzing the steps are shown above the arrows. GMD - GDP-mannose 4,6 dehydratase, is an $NADP^+$ dependent enzyme in which the $NADP^+$ is reduced and oxidized during the catalytic cycle. GFS – GDP-fucose synthetase (GDP-4-keto-6 deoxy-mannose 3,5 epimerase 4-reductase).

Figure 2: A) Stereo ribbon representation of GFS monomer showing bound $NADP^+$ as a ball-and-stick. The N-terminus of the protein is labeled, N-ter. The secondary structural elements are labeled, strands with numbers and helices with letters, proceeding from the N-terminus toward the C-terminus. NADP+ is shown in a ball and stick representation. B) Ribbon representation of the GFS dimer showing the extensive interface. The figure dimer is viewed looking down the two fold. One monomer is in red, the other in blue. Interacting strands and helices are

labeled as in 2A. The figures were made using MOLSCRIPT [49] and rendered using RASTER3D [50].

Figure 3: Stereo C-α trace of GFS, shown in blue, superimposed on GalE, shown in red. In each case the bound co-factor is shown as a ball-and-stick with the same color scheme as the protein. On GFS, every tenth Cα is shown as a ball and numbered where possible.

Figure 4: Quanta was used to superimpose *E. coli* UDP-galactose 4 epimerase (GalE) and *E. coli* GDP-fucose synthetase (coli_GFS) as shown in figure 3. The two sequences were then aligned based upon the structural alignment and the human GDP-fucose synthetase (human_GFS) amino acid sequence was aligned to this pair. Identical residues are boxed in red, homologous in grey, and residues shared between two of the three proteins are boxed in blue.

Figure 5: A) Stereo ball-and-stick representation of the bonding of $NADP^+$ to GFS. The protein is shown in dark green and the co-factor in blue. Water molecules are shown as red balls and potential hydrogen bonds shown as thin black lines. B) A close up view of the NADP(H) binding. The bound NADPH is shown with thick bonds and the bound $NADP^+$
in thin bonds.

Figure 6: A ball and stick representation of GDP-4keto 6deoxy mannose binding model. The proposed binding site residues are shown with dark bonds and the substrate/NADPH nicotinamide ring shown with light gray bonds.

Figure 7: The potential mechanism of the reduction (upper) and epimerization (lower) reactions catalyzed by GDP-fucose synthetase. Tyr136 plays the central role in donating a proton during reduction and stabilizing the negatively charged enediol during epimerization. This facilitates both reactions at a single active site. Ser107 assists along with interactions from Lys140 and the nicotinamide ribose (not shown). Alternatively Ser107 may function as part of a proton shuttle with Tyr136 as proposed for GalE [34].

Figure 8: A) Typical MIRAS electron density after modification with SOLOMON, contoured at 1.5σ. Part of the final refined GFS model is shown in density for reference. B) 2Fo-Fc electron density for NADP phased with the rigid body refined uncomplexed GFS model. The final refined model for $NADP^+$ is shown for reference. C) $2F_O$-F_C density for NADPH phased with the rigid body refined, uncomplexed, GFS model. The final refined NADPH coordinates are shown for reference.

Detailed Description of the Invention and Preferred Embodiments

Results and Discussion

GDP-fucose synthetase is a member of the short chain family of dehydrogenases-reductases.

The GFS monomer forms a roughly two domain structure that provides the enzyme with the ability to bind co-factor and substrate (Figure 2a). The NADP(H) binding domain is the larger of the two and contains a central six stranded β-sheet flanked by two sets of parallel α-helices, common to the family of NAD(P) binding proteins (reviewed in [16]). The second, predominantly C-terminal domain is smaller and is responsible for binding substrate. It extends away from the other domain and forms a globular cluster of three alpha-helices and two small beta-sheets.

The N-terminal domain begins with an alternating alpha/beta repeat forming the first five strands and four flanking helices labeled in figure 2a as 1-A-2-B-3-C-D-4-E-5. Residue Asn165 marks the first transition into the second, substrate binding domain, where it enters a short beta-strand (strand 6), a 12 residue loop, helix F, and two more strands (strands7 and 8). At that point the chain returns to the first domain forming helices G and H and the final strand of the central β-sheet, strand 9. The remaining residues of GFS form the bulk of the substrate binding domain and consist of the secondary structural elements 10-I-11-12-J-K terminating with a short piece of coil.

The structure of GFS reveals it to be a member of the short chain dehydrogenase/reductase (SDR) family of enzymes (reviewed in [15]). This family of enzymes catalyze diverse sets of reactions using a conserved core tertiary protein fold and a serine, tyrosine, lysine triad of catalytic residues. GalE belongs to the SDR family and forms its own branch with enzymes that catalyze dehydrogenations, dehydrations, and epimerizations and isomerization. The relationship between *E. coli* GFS, previously known as YEFB, and GalE has been previously noted [17] and GFS has been assigned to the GalE branch of SDRs based upon sequence homology. Consistent with this observation the structures of GFS and GalE are closely related. The overall sequence identity between GalE and GFS is 25%, resulting in structures with a RMS difference in 184 Cα positions of only 0.8Å (Figure 3). Whilst most of the secondary structural elements of the two enzymes superimpose well there are also some significant differences.

The first large difference occurs after the N-terminal strand-helix-strand in which GalE has a 22 residue insertion, forming an additional flanking helix and strand at the front of the molecule (see Figure 4 for amino acid alignment). This insertion provides residues in GalE which interact with the adenine ribose of NAD(H) [18] and would cause steric clashes if $NADP^+$ were to bind to GalE. In the absence of this loop, Arg36 of GFS directly hydrogen bonds with the C2' phosphate of NADP(H) and provides GFS with the ability to distinguish NAD(H) from NADP(H). The absence of this loop in GFS results in NADPH binding in a more solvent exposed arrangement, consistent with the observation that NADPH binds, then transfers the hydride, and is then released as $NADP^+$. In contrast GalE

does not release NAD^+ during the catalytic cycle and the nicotinamide dinucleotide is less solvent exposed.

For the next 150 residues of GFS there are only minor changes between the two protein in the positions of loops and flanking helices until His170 where there is a 6 residue insertion that extends GalE further into the solvent. Following this there is a helix in the substrate binding domain (helix F in GFS) that superimposes well with GalE and then two strands (corresponding to strands 7 and 11 in GFS), shown at the top of figure 3, that have both moved. These strands give the substrate binding region of GFS a more open, solvent exposed configuration and lack the "flap" in GalE that interacts with the substrate. From modeling of GDP-4-keto, 6-deoxy mannose binding to GFS (see below) some movement of residues within these loops may occur, as has been seen for other SDR enzymes [19]. The only remaining large difference between the two structures is an insertion of a helix from Ala228-Asn235 in GFS. This insertion is far from substrate or cofactor binding and therefore has unknown function.

In solution GFS exists as a dimer both from dynamic light scattering and size exclusion chromatography (data not shown). In the crystal lattice GFS exists as a crystallographic dimer and has an extensive monomer-monomer interface, burying 1530$Å^2$ of water accessible surface per monomer, as calculated with the CCP4 programs AREAMOL and RESAREA [20]. The core of the dimer interface is formed by a four helix bundle consisting of the flanking helices D and E interacting with themselves through a two fold rotation. This interface also includes some contacts between the loop Leu125-Leu129. The predominant interactions are between hydrophobic side chains on the long flanking helices along

with several hydrogen bonds at the periphery of the interface. This extensive interface presumably explains why the monomer is not observed in solution. Multimerization through a four helix bundle motif is a common feature in the SDR family with GalE [21, 22], 17 beta-hydroxysteroid dehydrogenase [23], and Dihydropteridine reductase [24] being typical examples of dimers formed this way.

NADP(H) binding

We obtained binary complexes with both $NADP^+$ or NADPH bound to GFS. $NADP^+$ lies against one face of the central beta-sheet with the N-terminal end of the first helix in GFS directed towards one of the adenine phosphoryl oxygens (Figures 2, 3, and 5a). $NADP^+$ binds in an extended conformation, such that it contacts almost every beta-strand and positions the nicotinamide ring in close proximity to the catalytic domain. The adenine and nicotinamide ribose conformations are C2' endo and C3' endo, respectively, with the nicotinamide ring in the *anti* conformation with respect to the ribose ring. The interactions made with the protein are a combination of direct and water mediated hydrogen bonds together with some hydrophobic interactions. The adenine ring packs between the side chain of Arg36 and the side chains of Leu41, Ala63 and Ile86. Arg36, which is disordered in the $NADP^+$ free structure, also makes hydrogen bonds with the ribose phosphoryl oxygens (Nε-OP3 and NH2-OP3 2.5Å and 2.4Å respectively). The only hydrogen bond to the adenine moiety is from the N6 to the OD1 of Asn40. One other phosphoryl oxygen also makes a water mediated hydrogen bond to the N of Arg36. The remaining water mediated hydrogen bond is between the adenine ribose O3 to the N of Arg12. The interactions with the phosphate groups are

similar to the characteristic NAD(P) binding domains of the dehydrogenases (Lesk, 1995). The turn between the end of the N-terminal strand and the N-terminal helix contains the characteristic GXXGXXG motif also observed in the structure of GalE. The phosphates lie within the helix dipole at the N-terminal end of the first helix and make hydrogen bonds with the N atoms of Met14 (2.8Å) and Val15 (2.8Å). The nicotinamide ribose hydroxyls make potential hydrogen bonds with the OH of Tyr136 (2.8Å), the Nε of Lys140 (3.0Å) and the carbonyl oxygen of Gly106 (2.3Å). The nicotinamide ring packs against Leu166 and makes potential hydrogen bonds with the OGs of Ser107 (2.7Å) and Ser108 (2.7Å) and the N of Ser108 (3.3Å). A comparison between the $NADP^+$ free and bound complexes shows that there are surprisingly few structural changes in GFS upon dinucleotide binding.

The alignment of *E. coli* and human GFS reveals that all residues involved in $NADP^+$ binding mentioned above are identical to or replaced with conservative substitutions in the human enzyme. The exception is Arg36 of the *E. coli* enzyme which is replaced by Phe40 in the human sequence. Arg36 coordinates the 2' phosphate group $NADP^+$, thereby allowing the enzyme to discriminate between $NADP^+$ and NAD^+. The inability of phenylalanine to make the necessary contacts allowing the enzyme to distinguish between $NADP^+$ and NAD^+, suggests that the local structures of the two enzymes differ in this area. At this time it we cannot say which residues in the human enzyme interact with the 2' phosphate group of $NADP^+$.

The structure of bound NADPH is superimposable on that of $NADP^+$ except for the nicotinamide ring, which rotates into the *syn* conformation relative

to the ribose ring (Figure 5b) and hydrogen bonds with phosphoryl oxygen. Inspection of the electron density (Figure 8c) revealed the expected slight puckering of the nicotinamide ring. As a consequence of this nicotinamide ring rotation, the hydrogen bonds with residues Ser107 and Ser108 are broken and two water molecules move into the site. One water molecule replaces the interactions made with the N7 and O7 and the other hydrogen bonds with Tyr136 OH and Ser107 OG.

NADPH binding in the *syn* confirmation allows transfer of the pro-*S* hydride (B-side) during catalysis. This accords with the known stereochemistry of the hydride transfer, (R. Kumar and G.-Y. Xu, personal communication). Transfer of the pro-*S* hydride is a general feature of SDR enzymes and NAD(P) has been shown to bind in the *syn* conformation in the structures of all the SDR enzymes solved to date [19, 22, 24-31]. In contrast, the product of the GDP-fucose synthetase reaction, NADP+, binds in the *anti* confirmation. It is conceivable that the different binding mode for substrate and product may help to account for the difference in affinity between the two and help promote product release. However the gain of H-bonds to the O7 and N7 of the nicotinamide ring in the binding of the product, $NADP^+$, relative to the substrate, NAPH, does not support this hypothesis. It seems more likely that the binding of $NADP^+$ in the *anti* conformation is an artifact of binding in the absence of the GDP-sugar substrate. The modeling described below suggests that the Ser107-Ser108 could move to interact with the mannose ring when substrate binds and that the *anti* conformation seen for $NADP^+$ is a consequence of an empty substrate binding site. UDP-glucose-4-epimerase also gave complexes with the nicotinamide ring bound in either *syn* or *anti*

confirmation depending upon the oxidation state of the cofactor, although in contrast to GFS the reduced cofactor was bound in the *anti* conformation. [18, 22]. However, in the structure of the ternary complex of GalE with UDP-sugar substrates, NADH bound in the *syn* conformation, the proper orientation to carry out hydride transfer [32. 33].

Substrate binding and the catalytic site.

Attempts to soak the GDP-4-keto, 6-deoxy mannose substrate or GDP into the crystals failed so a crude model of GDP-sugar binding was generated (Figure 6), based on the ternary complexes of GalE [32, 33, 34]. GDP-4-keto, 6-deoxy mannose was modeled in QUANTA and minimized with CHARM. The resulting structure was aligned with UDP- glucose in GalE (PDB accession 1KVU), then moved to optimize the hydrogen bond between the alpha phosphoryl oxygen and the N of Val180. Some adjustments of torsion angles within GDP-4-keto, 6-deoxy mannose were made to relieve some bad contacts and maximize van der Waals interacts. This model can be used to predict which residues may be important for substrate binding and catalysis. In the model, the Guanine ring of the GDP-sugar substrate lies in a hydrophobic pocket made by the side chains of Leu184, Val201, and Val180 and lies next to Trp202. In GalE this tryptophan is replaced by a phenylalanine which partially covers the bound substrate. When GDP-4-keto, 6-deoxy mannose binds to GFS this tryptophan may also move to partially bury the substrate. The N of Val180 hydrogen bonds to a guanosine phosphoryl oxygen which lies at the N-terminal end of helix Val180-Ala193. The model predicts that Lys283 and Arg209 may be involved in phosphate binding and that Ser107,

Ser108, Cys109 and Asn165 make interactions with the 4-keto sugar. The remaining side chain His179 is in proximity to act as the general acid or base during catalysis. The model also places the ketone oxygen within 4Å of the nicotinaimde ring, in close proximity and in the proper orientation for hydride transfer. The conserved catalytic triad, residues Ser107, Tyr136, and Lys140, occupy similar positions as in the GalE structure and are positioned to play a role in catalysis (see below).

Mechanisms of the reactions

A common theme in the reactions catalyzed the GalE and other SDR enzymes is the role played by the conserved Ser-Tyr-Lys. In the proposed mechanism, the pKa of the catalytic tyrosine is lowered via interactions with the positively charged lysine, the ribose hydroxyls of the nicotinamide, and potentially the catalytic serine [19, 22 ,23, 26 27, 34]. This allows the tyrosine to play the role of a general acid or base depending upon the reaction being catalyzed. The catalytic serine may also interact with the substrate stabilizing its conformation. This mechanism is supported by the structure of ternary complexes of GalE with NADH and UDP-sugars [18, 22, 32, 33] and mutagensis experiments with GalE [34, 35], as well as the structure of ternary complexes of other SDR enzymes [19, 26, 27] and mutagenesis of other SDR family members [36-40]. In GFS, Ser107, Tyr136, and Lys140 are properly positioned to play an analogous role in the epimerization and reductions reactions the enzyme catalyzes. In the GFS structure we find the distance between Nζ of Lys140 and the hydroxyl of Tyr136 (4.1Å) is too far to stabilize the negative charge on the tyrosine hydroxyl by hydrogen bond

interaction. Instead, as has been proposed for other SDR enzymes, Lys140 helps to stabilize the nicotinamide substrate in an active conformation through interactions with the ribose hydroxyls and may help lower the pKa of Tyr136 through electrostatic effects [19, 26, 27, 34].

In contrast to GalE and other SDR enzymes, GDP-fucose synthetase catalyzes two distinct sets of reactions, the epimerizations of C3 and C5 of the 4-keto, 6 deoxy-mannose ring and the NADPH dependent reduction at C4. The epimerizations at C3 and C5 differ from the epimerization reaction catalyzed by GalE, in that they do not involve the transient reduction and oxidation of an NAD^+ or $NADP^+$ cofactor. The epimerizations catalyzed by GFS most likely proceed through the enediol/enolate intermediate as first proposed by Ginsberg [41]. The same mechanism has been proposed for the epimerization reactions in the synthesis of related deoxy and dideoxy sugar-nucleotides (reviewed in [14, 42]).

In the epimerization catalyzed by GFS we propose that Tyr136, by virtue of its lowered pKa, plays the role of a general acid during catalysis. It transiently protonates the C4 oxygen, thereby stabilizing the enediol/enolate intermediate. The side chain of His179, as noted above, could fulfil the role of a general base in one of the reactions, abstracting a proton from C3 or C5 of the intermediate, followed by reprotonation from the opposite face of the sugar ring. Deprotonation of the C4 oxygen by Tyr136 acting as a general base completes the epimerization reaction. Lacking the structure of the ternary complex we cannot identify the other residues that function as active site acids or bases. This mechanism is consistent with the observed loss of the C3 proton during GFS catalyzed epimerization [10]

and with the ability of GFS to catalyze the epimerization reactions in the absence of NADPH and subsequent reduction at C4 (F. Sullivan unpublished data).

The other reaction catalyzed by GFS, the NADPH dependent reduction at C4 of the 4-keto, 6-deoxy-mannose ring, is more typical of reactions catalyzed by SDR enzymes. Here we propose that Tyr136, acts as a general acid and protonates the C4 oxygen in concert with hydride transfer to C4 from NADPH. Ser107 may play role in this reaction acting a proton shuttle or in stabilizing the conformation of the substrate in the active site, both of which have been suggested for other SDR enzymes [19, 26, 27, 34]. The common roles suggested for Tyr136 in the epimerization and reduction reactions are diagramed in Figure 7. It provides the mechanistic continuity between the distinct epimerazation and reduction reactions and suggests how they may be facilitated at the single active site in GFS. The details of the both the epimeraization and reduction reactions should be clarified by identification of a new crystal form of GFS which binds both the NADP(H) and GDP-sugar substrates and site directed mutagenesis of the implicated residues.

The residues in the substrate binding site are almost completely conserved between human and the *E. coli* sequences (Figure 4). The exception is Ser108 which is replaced with a conservative threonine mutation. Given the sequence similarity in the residues in the active sites of the human and *E. coli* enzymes, the *E. coli* structure may be a reasonable starting point to identify possible inhibitors of human GDP-fucose synthetase.

Both enzymes involved in GDP-fucose biosynthesis evolved from a common precursor.

Comparison of amino acid sequences reveals that the first enzyme in GDP-fucose biosynthesis, GDP-mannose 4,6 dehydratase, is as closely related to GDP-fucose synthetase (24% identity) as it is to UDP-glucose 4-epimerase (24% identity). GDP-mannose 4,6 dehydratase also contains the conserved Ser-Tyr-Lys catalytic triad. This suggests that all three enzymes have closely related structures and that both the enzymes involved in GDP-fucose biosynthesis evolved from a single ancestral gene. Additionally, it is interesting to note that the $NADP^+$ in GMD is transiently reduced and then reoxidized in the course of the reaction cycle, a role analogous to the one played by of NAD^+ in GalE. Both enzymes are known to bind their cofactors tightly during the catalytic cycle in order to prevent release of the transiently reduced nicotinamide [43]. Comparison of their sequences reveals that the loop that is thought to be responsible for the tight binding of cofactor in GalE, residues Leu33 - Phe54 (Figure 4), while absent in GFS, is present in GMD (data not shown). We predict that these residues also form a flap in GMD to provide additional interaction to keep the $NADP^+$ tightly bound during the catalytic cycle.

Biological implications

Fucose is found in the glycoconjugates of bacteria, plants and animals where it plays roles in maintaining structural integrity as well as in molecular recognition. Defects in GDP-fucose biosynthesis have been shown to affect nodulation in bacteria, stem development in plants and immune regulation in humans. GDP-fucose is synthesized from GDP-mannose by two enzymes, a $NADP^+$ dependent dehydratase and a dual function NADPH dependent epimerase-reductase, GDP-fucose synthetase. In this latter aspect biosynthesis of fucose differs from that of other deoxysugars which utilize separate epimerase and reductase enzymes.

Here we report the structure of *E. coli* GDP-fucose synthetase and binary complexes with $NADP^+$ and NADPH. This has allowed us to identify interactions involved in binding the NADPH substrate and to suggest the location of the binding site for the GDP-sugar substrate. Based upon these structures it appears that the enzyme contains a single active site that catalyzes both the epimerization and NADPH dependent reduction reactions. The residues in the active sites of the human and E. *coli* GDP-fucose synthetase are highly conserved. Thus the present structure of *E. coli* enzyme could serve as a starting point for the design of inhibitors of the human enzyme, which ultimately could lead to the design of immunosuppressants that act by blocking selectin mediated cell adhesion.

Material and Methods

Protein purification and crystallization

GFS protein was purified from an *E. coli* strain over-expressing the *E. coli fcl* gene, essentially as described by Sullivan et al. [9]. An additional step was added to the purification. The protein pool from the Heparin toyapearl step was made 1 M in $(NH_4)_2SO_4$ and loaded onto a Polypropyl A column (PolyLC). The column was eluted with a gradient from 1 to 0 M $(NH_4)_2SO_4$. The resulting protein was found to be monodisperse by light scattering analysis (DynaPro-801) and have a molecular weight consistent with a dimer. Similar results were obtained by gel filtration chromatography on a G3000 column (TosoHass). Crystals measuring 0.5x0.5x0.5mm were obtained within one week using the vapor diffusion hanging drop method. Hanging drops were set up by adding 10 ul of a 6 mg/mL protein solution in 10mM, pH 7.4 Tris buffer, 50 mM sodium chloride to 10 ul of the well solution consisting of 4.0M sodium formate.

Data collection and processing

Diffraction data were collected using a Raxis II detector mounted on an RU200 X-ray generator running at 50KV, 100mA, with the MSC/Yale focusing mirrors. All data collections were performed at 18°C with exposure times between 8 and 12 minutes per one degree oscillation. The data were reduced with DENZO/SCALEPACK [44] giving unit cell parameters of a=104.2Å and c=74.9Å and symmetry $P3_221$ or $P3_121$. The data are summarized in Table 1. The CCP4

suite of programs [20] were used for all further data processing leading up to heavy atom refinement.

MIRAS phasing

Initial attempts to solve the structure using molecular replacement with the homologous GalE structure as a search model failed. A similar attempt at molecular replacement by Tonetti et al. using data from similar crystals of GFS also was unsuccessful [45]. The structure was determined using three heavy atom derivatives. Crystals were soaked for 48 hr. in three different heavy metal salts, 5 mM gold potassium cyanide, 2 mM mercury acetate and 5 mM cadmium chloride, all dissolved in a 4.2M sodium formate crystal stabilization solution. The primary mercury acetate heavy atom position was determined by inspection of the Patterson function Harker sections and refined using MLPHARE [20]. One heavy atom site for the gold derivative and two sites for the cadmium were located with difference Fouriers. The space group was found to be $P3_221$ giving maps with good solvent protein boundaries and density that corresponded to many of the secondary structural elements of GalE. The gold and mercury heavy atom derivatives had single well occupied heavy atom sites close to Cys 249 in the final model, giving maps that were interpretable but with many main chain breaks. An additional heavy metal binding site was seen in the cadmium derivative. Heavy atom refinement in SHARP [46] revealed several minor sites for each derivative and a final figure of merit of 0.75 and 0.81 for acentric and centric reflections respectively. After density modification in SOLOMON [47] using a solvent content of 60%, the final

figure of merit for all reflections was 0.93. These maps were very high quality with no main breaks for the entire molecule (Figure 8a)

Model building and refinement

The model was built into the experimental maps using QUANTA (Molecular Simulations Inc.). Large pieces of GalE were used to assist with the model building by changing the side chain identities and moving residues and secondary structural elements. The resulting model had no breaks in the backbone and was refined using XPLOR positional, torsion angle dynamics, and B-factor refinement giving a final model with statistics shown in Table 2. The final model consists of residues Lys3 to Phe319 with the first and last two residues not visible in the electron density maps. The side chains of Arg36, Asp37, Arg55 and His174 are also disordered and were modeled as alanines in the final structure. The side chains of Arg36 and Asp37 became well ordered upon binding $NADP^+$ or NADPH and were therefore included in those complex models.

Obtaining NADP and NADPH bound complexes

The complex of GFS with $NADP^+$ was obtained by placing the crystals into 4.2M sodium formate, 1mM $NADP^+$ for 20 hours. The resulting complex was found to be isomorphous with cell parameters a=104.2Å and c=75.1Å. After rigid body refinement of the protein model in XPLOR [48] clear density was identified for the bound ligand in both $2F_O$-F_C and F_O-F_C electron density maps. A model of the complex was built using QUANTA and side chains were adjusted to fit the

new electron density. Refinement of the complex was performed using positional and B-factor refinement in XPLOR, giving a final model with statistics in Table 2.

The isomorphous complex with NADPH was produced by soaking existing crystals. A 3mM stock of NADPH was made in the 4.2M sodium formate solution and fully reduced by the addition of 100mM sodium borohydride. After 10 hours the crystal was placed into the resulting solution, soaked for 20 hours and then diffraction data were collected using methods described above. The crystal had cell parameters a=104.3Å and c=74.9Å and also gave clear electron density for NADPH in the resulting maps. This complex was refined using similar methods to the $NADP^+$ bound form.

Accession numbers

The coordinates of the apo enzyme structure, the $NADP^+$ complex, and NADPH complex have been deposited in the Protein Data Bank (entry codes 1GFS, 1FXS, and 1BSV).

References

1. Mergaert, P., Van Mantagu, M., & Holsters, M. (1997). The modulation gene nolK of Azorizobium caulinodans is involved in the formation of GDP-fucose from GDP-mannose. *FEBS Lett.* **409**, 312-316.

2. Bonin, C. P., Potter, I., Vanzin, G. F., & Reiter, W.-D. (1997). The MUR1 gene of Arabidopsis thaliana encodes and isoform of the GDP-D-mannose-4,6-Dehydratase, catalyzing the first step in the de novo synthesis of GDP-L-fucose. *Proc. Natl. Acad. Sci. USA* **94**, 2085-2090.

3. Kansas, G.S. (1996). Selectin and their ligands: current concepts and controversies. *Blood* **88**, 3259-3287.

4. Etzioni, A., *et al.*, & Gershoni-Baruch, R. (1992). Brief report: recurrent severe infections caused by a novel leukocyte adhesion deficiency. *N. Engl. J. Med.* **327**, 1789-1792.

5. Sturla, L., *et al.*, & Tonetti, M. (1998). Defective intracellular activity of GDP-D-mannose-4,6-dehydratase in leukocyte adhesion deficiency type II syndrome. *FEBS Lett* **429**, 274-278.

6. Karsan, A., *et al.*, & Harlan, J.M. (1998). Leukocyte Adhesion Deficiency Type II is a generalized defect of de novo GDP-fucose biosynthesis. Endothelial cell

fucosylation is not required for neutrophil rolling on human nonlymphoid endothelium. *J Clin Invest.* **101**, 2438-2445.

7. Broschat, K.O., Chang, S., & Serif, G. (1985). Purification and characterization of GDP-D-mannose 4,6-dehydratase from porcine thyroid *Eur. J. Biochem.* **153**, 397-401

8. Sturla, L., Bisso, A., Zanardi, D., Benatti, U., De Flora, A., & Tonetti, M. (1997). Expression, purification and characterization of GDP-D-mannose 4,6-dehydratase from *Escherichia coli. FEBS Lett.* **412**, 126-130.

9. Sullivan, F.X., et al., & Cumming, D.A. (1998) Molecular cloning of human GDP-mannose 4,6 dehydratase and reconstitution of GDP-fucose biosynthesis *in vitro. J. Biol. Chem.* **273**, 8193-8202.

10. Chang, S., Duerr, B., & Serif, G. (1988). An epimerase-reductase in L-fucose synthesis. *J. Biol. Chem.* **263**, 1693-1697

11. Tonetti, M., Sturla, L., Bisso, A., Benatti, U., & De Flora, A. (1996). Synthesis of GDP-L-fucose by the human FX protein. *J. Biol. Chem.* **271**, 27274-27279.

12. Stevenson, G., Anadrianopoulos, K., Hobbs, M., & Reeves, P.R. (1996) Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colonic acid. *J. Bacteriol.* **178**, 4885-4893.

13. Andrianopoulos, K., Wang, L., & Reeves, P. (1998). Identification of the fucose synthetase gene in the colanic acid gene cluster of *Escherichia coli* K-12. *J. Bacteriol.* **180**, 998-1001.

14. Liu, H.-W. & Thorson, J.S. (1994). Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. *Annu. Rev. Microbiol.* **48**, 223-256.

15. Jornvall, H., *et al.*, & Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34**, 6003-6013.

16. Lesk, A.M. (1995). NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* **5**, 775-783.

17. Persson, B., Krook, M., & Jornvall, H. (1995). Short-chain dehydrogenases/reductases. In *Enzymology and Molecular Biology of Carbonyl Metalbolism* , Vol. 5. (Weiner, H. *et al.* eds.), pp. 383-395, Plenum Press, New York.

18. Thoden, J.B., Frey, P.A., & Holden, H.M. (1996). Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated *from Escherichia coli. Biochemistry* **35**, 2557-2566.

19. Tanaka, N., Nonaka, T., Tanabe, T., Yoshimoto, T., Tsuru, D., & Mitsui, Y. (1996). Crystal structures of the binary and ternary complexes of 7 alpha-hydroxysteroid dehydrogenase *from Escherichia coli. Biochemistry* **35**, 7715-7730.

20. Collaborative Computational Project Number 4. (1994). The CCP4 suite: program for protein crystallography. *Acta Cryst.* D **50**, 760-763.

21. Bauer, A.J., Rayment, I., Frey, P.A., & Holden, H.M. (1992). The molecular structure of UDP-galactose 4-epimerase from *Escherichia coli* determined at 2.5 A resolution. *Proteins* **12**, 372-381.

22. Thoden, J.B., Frey, P.A., & Holden, H.M. (1996). High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol. *Protein Sci.* **5**, 2149-2161.

23. Ghosh, D., *et al.*, & Lin, S.X. (1995). Structure of human estrogenic 17 beta-hydroxysteroid dehydrogenase at 2.20 A resolution. *Structure* **3**, 503-513.

24. Varughese, K.I, Skinner, M.M., Whiteley, J.M., Matthews, D.A., & Xuong, N.H. (1992). Crystal structure of rat liver dihydropteridine reductase. *Proc. Natl. Acad. Sci.* U S A **89**, 6080-6084.

25. Ghosh, D., et al., & Orr, J.C. (1991). Three-dimensional structure of holo 3 alpha,20 beta-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family. *Proc. Natl. Acad. Sci. U S A* **88**, 10064-10068.

26. Breton, R., Housset, D., Mazza, C., & Fontecilla-Camps, J.C. (1996) The structure of a complex of human 17 beta-hydroxysteroid dehydrogenase with estradiol and $NADP^+$ identifies two principal targets for the design of inhibitors. *Structure* **4**, 905-915.

27. Tanaka, N., Nonaka, T., Nakanishi, M., Deyashiki, Y., Hara, A., & Mitsui, Y. (1996). Crystal structure of the ternary complex of mouse lung carbonyl reductase at 1.8 A resolution: the structural origin of coenzyme specificity in the short-chain dehydrogenase/reductase family *Structure* **4**, 33-45.

28. Rafferty, J.B., *et al.*, & Rice, D.W. (1995). Common themes in redox chemistry emerge from the X-ray structure of oilseed rape (*Brassica napus*) enoyl acyl carrier protein reductase. *Structure* **3**, 927-938.

29. Andersson, A., Jordan, D., Schneider, G., & Lindqvist, Y. (1996). Crystal structure of the ternary complex of 1,3,8-trihydroxynaphthalene reductase from *Magnaporthe grisea* with NADPH and an active-site inhibitor. *Structure* **4**, 1161-1170.

30. Nakajima, K., *et al.*, & Yamada, Y. (1998). Crystal structures of two tropinone reductases: different reaction stereospecificities in the same protein fold. *Proc. Natl. Acad Sci. U S A* **95**, 4876-4881.

31. Hulsmeyer M, *et al.*, & Schomburg, D. (1998). Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase from a PCB degrader at 2.0 A resolution *Protein Sci.* **7**, 1286-1293.

32. Thoden, J.B., Frey, P.A., & Holden, H.M. (1996). Molecular structure of the NADH/UDP-glucose abortive complex ofUDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism. *Biochemistry* **35**, 5137-5144.

33. Thoden, J.B., Hegeman, A.D., Wesenberg, G., Chapeau, M.C., Frey, P.A., & Holden, H.M. (1997). Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli. Biochemistry* **36**, 6294-6304.

34. Liu, Y., *et al.*, & Frey, P.A. (1997). Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from *Escherichia coli Biochemistry* **36**, 10675-10684.

35. Swanson, B.A. & Frey, P.A. (1993). Identification of lysine 153 as a functionally important residue in UDP-galactose 4-epimerase from Escherichia coli. *Biochemistry* **32**, 13231-13236.

36. Kiefer, P.M., Varughese, K.I., Su, Y., Xuong, N.H., Chang, C.F., Gupta, P., Bray, T., & Whiteley, J.M. (1996). Altered structural and mechanistic properties of mutant dihydropteridine reductases. *J. Biol. Chem.* **271**, 3437-3444.

37. Ensor, C.M. & Tai, H.H. (1994). Bacterial expression and site-directed mutagenesis of two critical residues (tyrosine-151 and lysine-155) of human placental NAD(+)-dependent 15-hydroxyprostaglandin dehydrogenase. *Biochim. Biophys. Acta* **1208**, 151-156.

38. Ensor, C.M. & Tai, H.H. (1996). Site-directed mutagenesis of the conserved serine 138 of human placental NAD+-dependent 15-hydroxyprostaglandin dehydrogenase to an alanine results in an inactive enzyme. *Biochem. Biophys. Res. Commun.* **220**, 330-333.

39. Obeid, J. & White, P.C. (1992). Tyr-179 and Lys-183 are essential for enzymatic activity of 11β-hydroxysteroid dehydrogenase. *Biochem. Biophys. Res. Commun.* **188**, 222-227.

40. Oppermann, U.C. et al. & Jornvall, H. (1997) Active site directed mutagenesis of 3 beta/17 beta-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions. *Biochemistry* **36**, 34-40.

41. Ginsberg, V. (1961). Studies on the biosynthesis of Guanosine Diphosphate L-fucose. *J. Biol. Chem.* **236**, 2389-2393.

42. Frey, P.A. (1987). Complex pyridine neuclotide-dependent transformations. *In Pyridine nucleotide coezymes: Chemical, biochemical, and medical aspects.* (Dolphin D, Poulson, R, Avromovic O, eds.), pp. 461-511, John Wiely and Sons, New York.

43. Oths, P.J., Mayer, R.M., & Floss, H.G. (1990). Stereochemistry and mechanism of the GDP-mannose dehydratase reaction. *Carbohydr Res* **198**, 91-100.

44. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

45. Tonetti, M., *et al.*, & Bolognesi, M. (1998). Preliminary cyrstallogrphic investigations of recombinant GDP-4-keto-6-deoxy-D mannose epimerase/reductase from *E. coli*. *Acta Cryst.* D **54**, 684-686.

46. La Fortelle, E. de, & Bricogne, G. (1997). Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enz.* **276** (Part B) 472-494.

47. Abrahams, J.P. & Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Cryst.* D **52**, 30-42.

48. Brunger, A.T. (1992). *X-PLOR. Version 3.1: a system for Crystallography and NMR*. Yale University Press, New Haven, CT.

49. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

50. Merrit, E.A. & Murphy, M.E.P. (1994). Raster3D version 2.0: a program for photorealistic molecular grahics. *Acta Cryst.* D **50**, 869-873.

All cited references are incoporated herein as if fully set forth.